

METRIC BASED ON AVERAGE DISTANCE BETWEEN SETS

OSAMU FUJITA

ABSTRACT. This paper presents a distance function between sets based on an average of distances between their elements. The distance function is a metric if the sets are non-empty finite subsets of a metric space. It can be applied to produce various metric spaces on collections of sets and will be useful for analyzing complex data sets in the fields of computer science and information science. Its generalizations to include the Hausdorff metric and extensions to infinite sets for treating fuzzy sets are also discussed.

1. INTRODUCTION

A metric defined in general topology [1, 2], based on a natural notion of distance between points, is generally extensible to distance between sets or more complex elements. The Hausdorff metric is such a typical one and practically used for image data analysis [3], but it has some problems. In the Euclidean metric on \mathbb{R} , for example, the Hausdorff distance between bounded subsets of \mathbb{R} depends only on their suprema or infima, no matter how the other elements of the sets are distributed within a certain range, which means it places importance on extremes and disregards the middle parts of the sets. This is a drawback because it is sensitive to noises, errors and outliers in analyzing real world data. There is a need to develop another metric that reflects the overall characteristics of elements of the sets.

In computer science, especially in the fields of pattern recognition, classification, information retrieval and artificial intelligence, it is important for data analysis to measure similarity or difference between data objects such as documents, images and signals. If the data objects can be represented by vectors, a conventional distance between vectors is a proper measure in their vector space. In practice, however, there are various data objects that should be dealt with in the form of collections of sets, probability distributions, graph structured data, or collections consisting of more complex data elements. To analyze these data objects, numerous distance-like functions have been developed [4], like the Mahalanobis distance and the Kullback-Leibler divergence, even though they do not necessarily satisfy symmetry and/or the triangle inequality.

As a true metric, besides the Hausdorff metric, there is another type of distance functions of sets, such as the Jaccard distance, based on the cardinality of the symmetric difference between sets or its variations. However, it measures only the size of the set difference, and takes no account of qualitative differences between individual elements. Thus, both metrics are insufficient to analyze informative data sets in which each element has its own specific meaning.

Key words and phrases. Metric, distance between sets, average distance, power mean, Hausdorff metric.

This paper presents a new distance function between sets based on an average distance. It takes all elements into account. It is a metric if the sets are non-empty finite subsets of a metric space, and includes the Jaccard distance as a special case. By using the power means [5], we obtain generalized forms that also include the Hausdorff metric. Extensions of the metric to hierarchical collections of infinite subsets will be useful for treating fuzzy sets and probability distributions.

2. PRELIMINARIES

The metric is extended to various types of generalized metrics. To avoid confusion in terminology, the following definition is used.

Definition 1. (Metric) Suppose X is a set and d is a function on $X \times X$ into \mathbb{R} . Then d is called a *metric* on X if it satisfies the following conditions, for all $a, b, c \in X$,

- M1:** $d(a, b) \geq 0$ (non-negativity),
- M2:** $d(a, a) = 0$,
- M3:** $d(a, b) = 0 \Rightarrow a = b$,
- M4:** $d(a, b) = d(b, a)$ (symmetry),
- M5:** $d(a, b) + d(b, c) \geq d(a, c)$ (triangle inequality).

The set X is called a *metric space* and denoted by (X, d) . The function d is called *distance function* or simply *distance*.

The metric is generalized by relaxing the conditions as follows:

- If d satisfies M1, M2, M4 and M5, then it is called a *pseudo-metric*.
- If d satisfies M1, M2, M3 and M5, then it is called a *quasi-metric*.
- If d satisfies M1, M2, M3 and M4, then it is called a *semi-metric*.

This terminology follows [1, 2], though the term “*semi-metric*” is sometimes referred to as a synonym of *pseudo-metric* [4].

A set-to-set distance is usually defined as follows (see, e.g., [6]): Let A and B be two non-empty subsets of X . For each $x \in X$, the distance from x to A , denoted by $\text{dist}(x, A)$, is defined by the equation

$$(2.1) \quad \text{dist}(x, A) = \inf\{d(x, a) \mid a \in A\}.$$

This is fundamental not only to the definitions of a boundary point and an open set in metric spaces but also to the generalization of a metric space to *approach space* [7]. Similarly, the distance from A to B can be straightforwardly defined by

$$(2.2) \quad \text{dist}(A, B) = \inf\{d(a, b) \mid a \in A, b \in B\}.$$

The function $\text{dist}()$ is neither a pseudo-metric nor a semi-metric. However, let $\mathcal{S}(X)$ be the collection of all non-empty closed bounded subsets of X . Then, for $A, B \in \mathcal{S}(X)$, the function $h(A, B)$ defined by

$$(2.3) \quad h(A, B) = \max\{\sup\{\text{dist}(b, A) \mid b \in B\}, \sup\{\text{dist}(a, B) \mid a \in A\}\}$$

is a metric on $\mathcal{S}(X)$, and h is called the *Hausdorff metric*. The collection $\mathcal{S}(X)$ topologized by the metric h is called a *hyperspace* in general topology.

In computer science, data sets are generally discrete and finite. A popular metric is the *Jaccard distance* (or *Tanimoto distance*, *Marczewski–Steinhaus distance* [4]) that is defined by

$$(2.4) \quad j(A, B) = \frac{|A \Delta B|}{|A \cup B|},$$

where $|A|$ is the cardinality of A , and \triangle denotes the symmetric difference: $A \triangle B = (A \setminus B) \cup (B \setminus A)$. In addition, $|A \triangle B|$ is also used as a metric.

In cluster analysis [8], the distance (2.2) is used as the *minimum distance* between data clusters for single-linkage clustering, and likewise the *maximum distance* is defined by replacing infimum with maximum for complete-linkage clustering. Moreover, the *group-average distance* (or *average distance*, *mean distance*) defined as $g(A, B)$ in the following is also typically used for hierarchical clustering. Although these three distance functions are not metrics, the group-average distance plays an important role in this paper.

Lemma 2. *Suppose (X, d) is a non-empty metric space. Let $\mathcal{S}(X)$ denote the collection of all non-empty finite subsets of X . For each A and B in $\mathcal{S}(X)$, define $g(A, B)$ on $\mathcal{S}(X) \times \mathcal{S}(X)$ to be the function*

$$(2.5) \quad g(A, B) = \frac{1}{|A||B|} \sum_{a \in A} \sum_{b \in B} d(a, b).$$

Then g satisfies the triangle inequality.

Proof. The triangle inequality for d yields $d(a, b) + d(b, c) - d(a, c) \geq 0$ for all $a, b, c \in X$. Then, for all $A, B, C \in \mathcal{S}(X)$, we have

$$(2.6) \quad \begin{aligned} & g(A, B) + g(B, C) - g(A, C) \\ &= \frac{1}{|A||B||C|} \sum_{a \in A} \sum_{b \in B} \sum_{c \in C} (d(a, b) + d(b, c) - d(a, c)) \geq 0. \end{aligned}$$

□

For ease of notation, let $s(A, B)$ be the sum of all pairwise distances between A and B such that

$$(2.7) \quad s(A, B) = \sum_{a \in A} \sum_{b \in B} d(a, b),$$

so that $g(A, B) = (|A||B|)^{-1}s(A, B)$. Since d is a metric, we have $s(A, B) \geq 0$, $s(A, B) = s(B, A)$, and $s(\{x\}, \{x\}) = 0$ for all $x \in X$. If $A = \emptyset$ or $B = \emptyset$, then $s(A, B) = 0$ due to the empty sum. If A and B are countable unions of disjoint sets, it can be decomposed as follows:

$$(2.8) \quad s\left(\bigcup_i^n A_i, \bigcup_j^m B_j\right) = \sum_i^n \sum_j^m s(A_i, B_j),$$

where $A_i \cap A_j = \emptyset = B_i \cap B_j$ for $i \neq j$. Furthermore, we define $t(A, B, C)$ by the following equation

$$(2.9) \quad t(A, B, C) = |C|s(A, B) + |A|s(B, C) - |B|s(A, C).$$

It follows from Lemma 2 that $t(A, B, C) \geq 0$ for $A, B, C \in \mathcal{S}(X)$, which is a shorthand notation of the triangle inequality (2.6).

3. METRIC BASED ON AVERAGE DISTANCE

Theorem 3. Suppose (X, d) is a non-empty metric space. Let $\mathcal{S}(X)$ denote the collection of all non-empty finite subsets of X . For each A and B in $\mathcal{S}(X)$, define $f(A, B)$ on $\mathcal{S}(X) \times \mathcal{S}(X)$ to be the function

$$(3.1) \quad f(A, B) = \frac{1}{|A \cup B| |A|} \sum_{a \in A} \sum_{b \in B \setminus A} d(a, b) + \frac{1}{|A \cup B| |B|} \sum_{a \in A \setminus B} \sum_{b \in B} d(a, b).$$

Then f is a metric on $\mathcal{S}(X)$.

Proof. The function f can be rewritten, using s in (2.7), as

$$f(A, B) = \frac{s(A, B \setminus A)}{|A \cup B| |A|} + \frac{s(A \setminus B, B)}{|A \cup B| |B|}.$$

It is non-negative and symmetric. If $A = B$, then $s(A, B \setminus A) = s(A, \emptyset) = 0$ and $s(A \setminus B, B) = s(\emptyset, B) = 0$, so that $f(A, B) = 0$. Conversely, if $f(A, B) = 0$, then $s(A, B \setminus A) = 0 = s(A \setminus B, B)$ for $A, B \in \mathcal{S}(X)$. This holds if, and only if, $B \setminus A = \emptyset = A \setminus B$, which implies $B \subseteq A$ and $A \subseteq B$. Then we have $f(A, B) = 0 \Leftrightarrow A = B$.

The triangle inequality is straightforwardly proved to be $f(A, B) + f(B, C) - f(A, C) \geq 0$ by showing that the left-hand terms are transformed into the sum of non-negative terms of s and t in (2.9). Let $A \cup B \cup C$ be decomposed into five disjoint partitions: $\alpha = A \setminus (B \cup C)$, $\beta = B \setminus (A \cup C)$, $\gamma = C \setminus (A \cup B)$, $\zeta = A \cap C \setminus B$, and $\theta = B \setminus \beta = B \cap (A \cup C)$. Then we have

$$\begin{aligned} & |A| |B| |C| |A \cup B| |B \cup C| |A \cup C| (f(A, B) + f(B, C) - f(A, C)) \\ &= |B| |C| (|\theta \cup C| t(A, B \setminus A, \gamma) + |\alpha| t(A, \beta, \gamma) + |B \cup \zeta| t(A, \beta, C \setminus A)) \\ &\quad + |A| |B| (|A \cup \theta| t(\alpha, B \setminus C, C) + |\gamma| t(\alpha, \beta, C) + |B \cup \zeta| t(A \setminus C, \beta, C)) \\ &\quad + |A| |C| (|A \setminus B| t(\alpha, B, C \setminus B) + |B| t(\alpha, \theta, \gamma) + |C \setminus B| t(A \setminus B, B, \gamma)) \\ &\quad + |A| |C| |\theta| (t(\alpha, B, \gamma) + t(\alpha, B, \zeta) + t(\zeta, B, \gamma)) \\ &\quad + 2 |A| |C| (|\theta \cup C| |A \cup \theta| + |\beta| |A \cup C|) s(B, \zeta) \\ &\quad + 2 |A| |B| |C| |B \cup \zeta| s(\beta, A \cap C) \geq 0. \end{aligned}$$

The details are given in Appendix A. □

The function f in (3.1) can be rewritten, using g in (2.5), as

$$(3.2) \quad f(A, B) = \frac{|B \setminus A|}{|A \cup B|} g(A, B \setminus A) + \frac{|A \setminus B|}{|A \cup B|} g(A \setminus B, B).$$

In $(\mathcal{S}(X), f)$, for all $a, b \in X$, we have $f(\{a\}, \{b\}) = d(a, b)$ so that $\{\{x\} \mid x \in X\}$ is an isometric copy of X . If $A \cap B = \emptyset$, then $f(A, B) = g(A, B)$. If d is a pseudo-metric, then so is f .

Example 4. If d is the discrete metric, where $d(x, y) = 0$ if $x = y$ and $d(x, y) = 1$ otherwise, then $f(A, B)$ is equal to the Jaccard distance (2.4).

Corollary 5. Suppose (X, d) is a non-empty metric space. Let $\mathcal{S}(X)$ denote the collection of all non-empty finite subsets of X . For each A and B in $\mathcal{S}(X)$, define $e(A, B)$ on $\mathcal{S}(X) \times \mathcal{S}(X)$ to be the function

$$e(A, B) = \frac{1}{|A| |B|} \left(\sum_{a \in A} \sum_{b \in B} d(a, b) - \sum_{a \in A \cap B} \sum_{b \in A \cap B} d(a, b) \right).$$

Then e is a semi-metric on $\mathcal{S}(X)$.

Proof. Let $e(A, B)$ be rewritten as

$$e(A, B) = \frac{1}{|A||B|} (s(A \setminus B, B \setminus A) + s(A \cap B, B \setminus A) + s(A \setminus B, A \cap B)).$$

In a similar manner to the proof of Theorem 3, it can be proved that the conditions from M1 to M4, except for M5 (triangle inequality), are satisfied. \square

It is noted that the triangle inequality $e(A, B) + e(B, C) \geq e(A, C)$ holds if $|\delta||\varepsilon||\eta| = 0$, where $\delta = A \cap B \setminus C$, $\varepsilon = B \cap C \setminus A$ and $\eta = A \cap B \cap C$. Otherwise, for example, if $A = \delta \cup \eta$, $B = \delta \cup \eta \cup \varepsilon$ and $C = \eta \cup \varepsilon$ for non-empty δ , ε and η , then we have

$$\begin{aligned} & |A||B||C| (e(A, B) + e(B, C) - e(A, C)) \\ &= |\delta|s(\delta, \varepsilon) - |\varepsilon|s(\delta, \eta) - |\delta|s(\eta, \varepsilon) = -t(\delta, \eta, \varepsilon) \leq 0, \end{aligned}$$

so that the condition M5 is not generally satisfied.

4. EXTENSIONS

This section discusses future directions for generalization of the average distance based on the power mean and extensions to metrics on collections of infinite sets.

4.1. Generalization based on the power mean. The distance function (3.1) can be unified with the Hausdorff metric for finite sets by using the power mean. To simplify expressions, we use the following notation. Let $M_p^{(i)}(x \in A, \psi, w)$ be an extended weighted-power-mean of $\psi(x)$ such that

$$(4.1) \quad M_p^{(1)}(x \in A, \psi, w) = \left(\frac{1}{\sum_{x \in A} w(x)} \sum_{x \in A} w(x) (\psi(x))^p \right)^{1/p},$$

and its variation using the exponential transform of ψ ,

$$(4.2) \quad M_p^{(0)}(x \in A, \psi, w) = \frac{1}{p} \ln \left(\frac{1}{\sum_{x \in A} w(x)} \sum_{x \in A} w(x) \exp(p\psi(x)) \right),$$

where $i \in \{0, 1\}$ indicates one of the two types (4.1) and (4.2), p is an extended real number, ψ is a non-negative function of $x \in A$, and w is a weight such that $w(x) \in (0, 1]$ for each x and $\sum_{x \in A} w(x) > 0$. In addition, let $M_p^{(i)}(x \in A, \psi)$ denote the abbreviation of the equal weight case $M_p^{(i)}(x \in A, \psi, 1_A(x))$, where $1_A(x)$ is the indicator function defined by $1_A(x) = 1$ for $x \in A$ and $1_A(x) = 0$ for $x \notin A$. If there exists $x \in A$ such that $\psi(x) = 0$ for $p < 0$ in (4.1), then we define $M_p^{(1)} = 0$, which is consistent with taking the limit $\psi(x) \rightarrow 0^+$, though such a case is undefined in the conventional power mean to avoid division by zero.

The power mean includes various types of means [5], which are parameterized by p . By taking limits also for $p = 0, \pm\infty$, we have the following:

$$\begin{aligned} M_i^{(i)}(x \in A, \psi(x), 1_A(x)) &= M_i^{(i)}(x \in A, \psi(x)) = \frac{1}{|A|} \sum_{x \in A} \psi(x), \\ M_\infty^{(i)}(x \in A, \psi(x), w(x)) &= M_\infty^{(i)}(x \in A, \psi(x)) = \max\{\psi(x) \mid x \in A\}, \\ M_{-\infty}^{(i)}(x \in A, \psi(x), w(x)) &= M_{-\infty}^{(i)}(x \in A, \psi(x)) = \min\{\psi(x) \mid x \in A\}. \end{aligned}$$

Both $M_1^{(1)}$ and $M_0^{(0)}$ give the same arithmetic mean, whereas $M_0^{(1)}$ gives the geometric mean, and neither $M_\infty^{(i)}$ nor $M_{-\infty}^{(i)}$ depends on $w(x)$.

There are some forms of function composition of $M_p^{(i)}$ that include the distance function (3.1) as a special case, for example, as follows:

$$(4.3) \quad u_{p,q}^{(i,j)}(A, B) = M_p^{(i)} \left(x \in A \cup B, \left(1_{B \setminus A}(x) M_q^{(j)}(y \in A, d(x, y)) \right. \right. \\ \left. \left. + 1_{A \setminus B}(x) M_q^{(j)}(y \in B, d(x, y)) \right) \right),$$

$$(4.4) \quad v_{r,p,q}^{(k,i,j)}(A, B) \\ = M_r^{(k)} \left(S \in \{A, B\}, M_p^{(i)} \left(x \in A \cup B, 1_{S^c}(x) M_q^{(j)}(y \in S, d(x, y)) \right) \right),$$

where $i, j, k \in \{0, 1\}$ and S^c is the complement of S .

In addition, let w be extended to $w \in [0, 1]$ to include zero, though a weight for at least one summand must still be positive. Furthermore, it is assumed that $0 \cdot \infty = 0$, in order to ensure $0 \cdot 0^p = 0$ for $p < 0$ in $M_p^{(1)}$, so that the zero-weight can be used for excluding terms from averaging even if $\psi(x) = 0$ (i.e., distance zero) in the terms. Then, the function (4.3) can be simply expressed as

$$(4.5) \quad u_{p,q}^{(i,j)}(A, B) = M_p^{(i)} \left(x \in A \cup B, M_q^{(j)}(y \in A \cup B, d(x, y), w(x, y)) \right),$$

by using the weight function defined by

$$w(x, y) = [x \in A \setminus B][y \in B] + [x \in A \cap B][x = y] + [x \in B \setminus A][y \in A] \\ = 1_{A \setminus B}(x) 1_B(y) + 1_{A \cap B}(x)[x = y] + 1_{B \setminus A}(x) 1_A(y),$$

where $[\cdot]$ denotes the Iverson bracket, that is a quantity defined to be 1 whenever the statement within the brackets is true, and 0 otherwise.

The distance function f in (3.1) and the Hausdorff metric (2.3) are expressed by

$$f(A, B) = u_{i,j}^{(i,j)}(A, B) = 2 v_{k,i,j}^{(k,i,j)}(A, B), \\ h(A, B) = u_{\infty,-\infty}^{(i,j)}(A, B) = v_{\infty,\infty,-\infty}^{(k,i,j)}(A, B),$$

respectively, where $i, j, k \in \{0, 1\}$.

Although it is unclear, at present, what conditions on the parameters i, j, k, p, q , and r are necessary for (4.3) and (4.4) to be metrics, these generalized forms are capable of generating various distance functions in fact as follows:

Example 6. The exponential types $u_{p,q}^{(0,0)}(A, B)$ and $v_{r,p,q}^{(0,0,0)}(A, B)$ are written as

$$u_{p,q}^{(0,0)}(A, B) = \frac{1}{p} \ln \left(\frac{1}{|A \cup B|} \sum_{x \in A \cup B} \left(\frac{1_{B \setminus A}(x)}{|A|} \sum_{y \in A} e^{qd(x,y)} \right. \right. \\ \left. \left. + 1_{A \cap B}(x) + \frac{1_{A \setminus B}(x)}{|B|} \sum_{y \in B} e^{qd(x,y)} \right)^{p/q} \right),$$

$$v_{r,p,q}^{(0,0,0)}(A,B) = \frac{1}{r} \ln \left(\frac{1}{2} \left(\frac{1}{|A \cup B|} \sum_{x \in B \setminus A} \left(\frac{1}{|A|} \sum_{y \in A} e^{qd(x,y)} \right)^{p/q} + \frac{|A|}{|A \cup B|} \right)^{r/p} \right. \\ \left. + \frac{1}{2} \left(\frac{1}{|A \cup B|} \sum_{x \in A \setminus B} \left(\frac{1}{|B|} \sum_{y \in B} e^{qd(x,y)} \right)^{p/q} + \frac{|B|}{|A \cup B|} \right)^{r/p} \right).$$

If d is the discrete metric multiplied by a positive constant λ , then we have

$$(4.6) \quad u_{p,q}^{(0,0)}(A,B) = \frac{1}{p} \ln \left(\frac{e^{p\lambda} |A \triangle B| + |A \cap B|}{|A \cup B|} \right),$$

$$(4.7) \quad v_{0,p,q}^{(0,0,0)}(A,B) = \frac{1}{2p} \ln \left(\left(\frac{e^{p\lambda} |B \setminus A| + |A|}{|A \cup B|} \right) \left(\frac{e^{p\lambda} |A \setminus B| + |B|}{|A \cup B|} \right) \right).$$

The functions (4.6) and (4.7) are metrics for $p > 0$ and $p < 0$, respectively. The proofs of each triangle inequality are outlined in Appendix B. If $p = 0$, then it is the same situation as Example 4. By taking the limit for $p = 0$, we can see that both functions are equal to the Jaccard distance (2.4) except for the coefficient λ .

4.2. Hierarchical metric spaces. Suppose that (X, d) is a metric space and $\mathcal{S}(X)$ is the collection of all non-empty finite subsets of X . Let k be a non-negative integer, let $\mathcal{S}_{k+1}(X)$ denote the collection of all non-empty finite subsets of $\mathcal{S}_k(X)$, and let f_k be a metric on $\mathcal{S}_k(X)$, where $\mathcal{S}_1(X)$, $\mathcal{S}_0(X)$, f_1 and f_0 correspond to, respectively, $\mathcal{S}(X)$, X , f , and d in Theorem 3. For $k > 1$, in much the same way, for each \mathcal{A} and \mathcal{B} in $\mathcal{S}_k(X)$, the function $f_k(\mathcal{A}, \mathcal{B})$ can be defined by

$$(4.8) \quad f_k(\mathcal{A}, \mathcal{B}) = \frac{\sum_{A \in \mathcal{A}} \sum_{B \in \mathcal{B} \setminus \mathcal{A}} f_{k-1}(A, B)}{|\mathcal{A} \cup \mathcal{B}| |\mathcal{A}|} + \frac{\sum_{A \in \mathcal{A} \setminus \mathcal{B}} \sum_{B \in \mathcal{B}} f_{k-1}(A, B)}{|\mathcal{A} \cup \mathcal{B}| |\mathcal{B}|},$$

which generates a metric space $(\mathcal{S}_k(X), f_k)$ based on $(\mathcal{S}_{k-1}(X), f_{k-1})$. This metric will be useful for constructing hierarchical hyperspaces.

4.3. Duality. There is a kind of duality between sets and elements with respect to their distance functions. For example, we can define the functions D and d symmetrically as follows:

$$(4.9) \quad D(A, B) = |\{a \mid a \in A\} \triangle \{b \mid b \in B\}| = |A \triangle B|,$$

$$(4.10) \quad d(a, b) = |\{A \mid a \in A\} \triangle \{B \mid b \in B\}| = |\mathcal{C}(a) \triangle \mathcal{C}(b)|,$$

where $\mathcal{C}(a) = \{A \mid a \in A\}$. The set-to-set distance D in (4.9) is a metric due to the axiom of extensionality, and the element-to-element distance d in (4.10) is a pseudo-metric.

According to Theorem 3, D can be defined by f in (3.1), instead of (4.9), so that we have

$$D(A, B) = f(A, B), \\ d(a, b) = |\mathcal{C}(a) \triangle \mathcal{C}(b)| = f \left(\bigcap \mathcal{C}(a), \bigcap \mathcal{C}(b) \right).$$

In this case, D depends on d so that D is also a pseudo-metric. Furthermore, in the situation of Section 4.2, let (X, d) be a metric space, let $\mathcal{S}_1(X)$ be the collection

of all non-empty finite subsets of X , and let $\mathcal{C}(a) = \{A \in \mathcal{S}_1(X) \mid a \in A\}$ be an element of $\mathcal{S}_2(X)$. If d is the discrete metric, then we have

$$\begin{aligned} D(A, B) &= f_1(A, B) = j(A, B), \\ f_2(\mathcal{C}(a), \mathcal{C}(b)) &= \kappa d(a, b), \end{aligned}$$

where κ is a certain positive real number such that $\kappa < 1$. If there exists d such that $d(a, b) = f_2(\mathcal{C}(a), \mathcal{C}(b))$, then the isometric copy of (X, d) is contained in $(\mathcal{S}_2(X), f_2)$ and d can be regarded as the function of D . In general, there possibly exist D and d that are formally expressed by

$$(4.11) \quad D(A, B) = F(\{d(a, b) \mid a \in A, b \in B\}),$$

$$(4.12) \quad d(a, b) = G(\{D(A, B) \mid a \in A, b \in B\}),$$

where F and G may be such a generalized function given in (4.5). It is interesting to consider whether F and G really exist and what features they have. In numerical analysis, D and d that are consistent with each other will be obtained by the iterative computation of (4.11) for all $A, B \in \mathcal{S}_1(X)$ and (4.12) for all $a, b \in X$, starting with an initial metric space (X, d) , if each converges to a non-trivial function.

4.4. Generalized metrics. The group average distance $g(A, B)$ in (2.5) can be regarded as a generalized metric that satisfies conditions M1 (non-negativity), M4 (symmetry) and M5 (triangle inequality) in Definition 1. In conventional topology, there has been no such generalization by dropping both conditions M2 and M3, which are usually combined together into the single axiom $d(a, b) = 0 \Leftrightarrow a = b$ (identity, reflexivity, or coincidence). Although M3 can be dropped for the pseudo-metric so as to allow $d(a, b) = 0$ for $a \neq b$, the self-distance $d(a, a) = 0$ (M2) seems to be indispensable in point-set topology where the element is a point having no size. An exception is a *partial metric* [9] that is defined to satisfy M1, M3, M4 and, instead of M5, the following partial metric triangularity,

$$(4.13) \quad d(a, b) + d(b, c) \geq d(a, c) + d(b, b).$$

In computer science or information science, the element of data sets is not merely a simple point. It may have rich contents inside. Some elements may have internal structures which cause non-zero self-distance, and some elements may have different properties each other, even though they are indiscernible from a metric point of view. The concept of distance can be used for measuring not only the difference between objects but also the cost of moving or the energy of transition between states. This is the reason why the generalization toward non-zero self-distance is worth considering. If the triangle inequality holds, it provides an upper and lower bound for them. The group average distance $g(A, B)$ can be such a typical one, and that it is simpler and more natural than the metric $f(A, B)$ in (3.1).

Incidentally, the function $g(A, B)$ is not a partial metric because it does not satisfy (4.13). On the other hand, $f(A, B)$ gives an approach to an instance of partial metrics from a special case of (4.7) in Example 6. By taking the limit as $\lambda \rightarrow \infty$ for $p < 0$ and multiplying a positive constant, for non-empty finite sets A and B , we have the following metric,

$$D_\nu(A, B) = \log |A \cup B| - \nu \log(|A| |B|),$$

where $\nu = 1/2$. Its triangle inequality is equivalent to $|A \cup B| |B \cup C| \geq |A \cup C| |B|$. This suggests, for $\nu \in [0, 1/2)$, D_ν is a partial metric on a collection of non-empty finite sets.

4.5. Extension to infinite sets. If $\mathcal{S}(X)$ is the collection of all non-null measurable subsets of (X, d) , and d is Lebesgue integrable on each element of $\mathcal{S}(X)$, then the group average distance $g(A, B)$, for $A, B \in \mathcal{S}(X)$, can be defined by

$$(4.14) \quad g(A, B) = \frac{1}{\mu(A)\mu(B)} \int_A \left(\int_B d(x, y) d\mu(y) \right) d\mu(x)$$

where $x \in A$, $y \in B$, and μ is a measure on X , and then the distance function (3.2) can be extended to

$$(4.15) \quad f(A, B) = \frac{\mu(B \setminus A)}{\mu(A \cup B)} g(A, B \setminus A) + \frac{\mu(A \setminus B)}{\mu(A \cup B)} g(A \setminus B, B).$$

If d is the discrete metric, then (4.15) is equal to the *Steinhaus distance* [4].

Example 7. Let (\mathbb{R}, d) be a metric space and let $d(x, y) = |x - y|$. For two intervals A and B , the distance function (4.15) can be expressed as

$$f(A, B) = \frac{|\sup(A) - \sup(B)| + |\inf(A) - \inf(B)|}{2} - \frac{|\sup(A) - \sup(B)| |\inf(A) - \inf(B)|}{\sup(A \cup B) - \inf(A \cup B)} [(A \subset B) \vee (A \supset B)]$$

If $A \not\subset B$ and $A \not\supset B$ (i.e., $[(A \subset B) \vee (A \supset B)] = 0$), then $f(A, B)$ is equal to the distance between the centers of A and B . This is consistent with an intuitive notion of the distance between balls in this (\mathbb{R}, d) .

If $\mathcal{S}(X)$ is the collection of all non-empty, countably infinite subsets (measure-zero sets) of X , then $g(A, B)$ and $f(A, B)$ should be defined by taking limits in (2.5) and (3.2), provided that both have definite values. In order to determine the average distance, we have to define a proper condition, which should be said to be “averageable”. The average distance will strongly depend on accumulation points in A and B , and it will require additional assumptions on the difference of the strength between the accumulation points. This requirement is closely related to a “relative measure” that is needed to obtain the ratio of the cardinality of an infinite set to the cardinality of its superset in (3.2). In conventional measure theory, however, any set of cardinality \aleph_0 is a null set having measure zero so that both counting measure and Lebesgue measure are useless for computing the ratio. It is necessary to use another measure. A feasible solution is discussed in the following section.

4.6. Estimation by sampling. In application to computational data analysis, statistical estimation by sampling is very useful for obtaining the approximate value of $f(A, B)$ when the size of the sets is very large. According to the law of large numbers, if enough sample elements are selected randomly, an average generated by those samples should approximate the average of the total population. The procedure is as follows:

- (1) Choose a superset P of $A \cup B$ as a population such that $P \supseteq A \cup B$.
- (2) Select a finite subset S of P as a sample obtained by random sampling.
- (3) Let $S_A = S \cap A$ and $S_B = S \cap B$. Then, compute $f(S_A, S_B)$ for approximation of $f(A, B)$.

The sampling process and its randomness are crucial for efficiently estimating a good approximation. Some useful hints could be found in various sampling techniques developed for Monte Carlo methods [10]. In most cases, sampling error is expected to decrease as the sample size increases, except for situations where the distribution of d has no mean (e.g., Cauchy distribution).

The notion of sampling suggests an intuitive measure to define a relative measure on a σ -algebra over a set X , which could be called “sample counting measure”. Suppose A and B are subsets of X . Let $\rho(A : B)$ be the ratio of the cardinality of A to the cardinality of B , let P be a superset of $A \cup B$, and let S_n be a non-empty finite subset of P such that $S_n = \bigcup_{i=1}^n Y_i$ where Y_i is the i -th non-empty sample randomly selected from P . Then, the ratio $\rho(A : B)$ can be determined by taking a limit of n as it approaches to ∞ as follows:

$$\rho(A : B) = \lim_{n \rightarrow \infty} \frac{|S_n \cap A|}{|S_n \cap B|},$$

if there exist such a limit and a random choice function that performs random sampling. Otherwise, instead of random sampling, systematic sampling could be available if the elements of X are supposed to be distributed with uniform density in its measurable metric space. For example, suppose there exists a finite partition of P where every part has an almost equal diameter. It seems better for S_n to have exactly one element with each of the parts.

4.7. Metrics for fuzzy sets and probability distributions. A fuzzy set can be represented by a collection of crisp sets so that the distance between fuzzy sets can be defined by the distance between the collections of such crisp sets. Let A be a fuzzy set: $A = \{(x, m_A(x)) \mid x \in X\}$, where $m_A(x)$ is a membership function, and let A_α be a crisp set called an α -level set [11] such that $A_\alpha = \{x \in X \mid m_A(x) \geq \alpha\}$. Then, A can be represented by the following set of ordered pairs: $\mathcal{C}(A) = \{(A_\alpha, \alpha) \mid \alpha \in (0, 1]\}$. The distance between two fuzzy sets A and B can be defined by $f_2(\mathcal{C}(A), \mathcal{C}(B))$, where there may be various ways to treat α . This notion is also applicable to the distance between probability distributions, where probability density functions are used instead of the membership function.

5. CONCLUDING REMARKS

We have found that, for a metric space (X, d) , there exists a distance function between non-empty finite subsets of X that is a metric based on the average distance of d . The distance function (3.1) in Theorem 3 is the most typical one, which includes the Jaccard distance as a special case where d is the discrete metric. Its extensions based on the power mean will be useful to develop generalized forms that also include the Hausdorff metric and the other various distance functions. Furthermore, the extensions to infinite subsets of X will provide metrics for measuring dissimilarity of fuzzy sets and probability distributions.

Appendix A. Triangle Inequality in Theorem 3

The triangle inequality for

$$f(A, B) = (|A \cup B| |A|)^{-1} s(A, B \setminus A) + (|A \cup B| |B|)^{-1} s(A \setminus B, B)$$

can be proved by showing the following inequality:

$$|A| |B| |C| |A \cup B| |B \cup C| |A \cup C| (f(A, B) + f(B, C) - f(A, C)) \geq 0.$$

Let $A \cup B \cup C$ be decomposed into the following seven disjoint sets:

$$\begin{aligned}\alpha &= A \setminus (B \cup C), & \beta &= B \setminus (A \cup C), & \gamma &= C \setminus (A \cup B), \\ \delta &= A \cap B \setminus C, & \varepsilon &= B \cap C \setminus A, & \zeta &= C \cap A \setminus B, & \eta &= A \cap B \cap C,\end{aligned}$$

and let $\theta = B \setminus \beta = B \cap (A \cup C)$, so that we have

$$\begin{aligned}A &= \alpha \cup \delta \cup \zeta \cup \eta, & |A| &= |\alpha| + |\delta| + |\zeta| + |\eta|, \\ B &= \beta \cup \delta \cup \varepsilon \cup \eta, & |B| &= |\beta| + |\delta| + |\varepsilon| + |\eta|, \\ C &= \gamma \cup \varepsilon \cup \zeta \cup \eta, & |C| &= |\gamma| + |\varepsilon| + |\zeta| + |\eta|, \\ \theta &= \delta \cup \varepsilon \cup \eta, & |\theta| &= |\delta| + |\varepsilon| + |\eta|.\end{aligned}$$

Taking account of (2.8) and (2.9), we have

$$\begin{aligned}& |A| |B| |C| |A \cup B| |B \cup C| |A \cup C| (f(A, B) + f(B, C) - f(A, C)) \\&= |B| |C| |B \cup C| |A \cup C| s(A, B \setminus A) + |A| |C| |B \cup C| |A \cup C| s(A \setminus B, B) \\&\quad + |A| |C| |A \cup B| |A \cup C| s(B, C \setminus B) + |A| |B| |A \cup B| |A \cup C| s(B \setminus C, C) \\&\quad - |B| |C| |A \cup B| |B \cup C| s(A, C \setminus A) - |A| |B| |A \cup B| |B \cup C| s(A \setminus C, C) \\&= |B| |C| (|\gamma| |\delta \cup C| s(A, B \setminus A) + (|\gamma| |\alpha| + |B \cup \zeta| |C \setminus A|) s(A, \beta) \\&\quad + |B \cup \zeta| |A| (s(A \setminus C, \beta) + s(A \cap C, \beta)) + (|\beta| |\gamma| + |B \cup C| |A \cup \varepsilon|) s(A, \varepsilon)) \\&\quad + |A| |C| ((|\gamma| |A \cup C| + |\zeta| |\gamma| + |\zeta| |A \cup \varepsilon|) s(\alpha, B) \\&\quad + |B| |A \cup \varepsilon| (s(\alpha, B \setminus C) + s(\alpha, B \cap C)) + |B| |\gamma| (s(\alpha, \beta) + s(\alpha, \theta)) \\&\quad + (|\delta \cup C| |\gamma| + |\delta \cup C| |A \cup \varepsilon| + |\beta| |A \cup C|) s(\zeta, B)) \\&\quad + |A| |C| ((|\alpha| |A \cup C| + |\zeta| |\alpha| + |\zeta| |\delta \cup C|) s(B, \gamma) \\&\quad + |B| |\delta \cup C| (s(B \setminus A, \gamma) + s(A \cap B, \gamma)) + |B| |\alpha| (s(\beta, \gamma) + s(\theta, \gamma)) \\&\quad + (|A \cup \varepsilon| |\alpha| + |A \cup \varepsilon| |\delta \cup C| + |\beta| |A \cup C|) s(B, \zeta)) \\&\quad + |A| |B| (|\alpha| |A \cup \varepsilon| s(B \setminus C, C) + (|\alpha| |\gamma| + |\zeta \cup B| |A \setminus C|) s(\beta, C) \\&\quad + |\zeta \cup B| |C| (s(\beta, A \cap C) + s(\beta, C \setminus A)) + (|\beta| |\alpha| + |A \cup B| |\delta \cup C|) s(\delta, C)) \\&\quad - |B| |C| (|\zeta \cup B| |\beta| s(A, C \setminus A) + (|\alpha| |\beta| + |B \setminus A| |\delta \cup C|) s(A, \gamma) \\&\quad + |A| |\delta \cup C| (s(\alpha \cup \zeta, \gamma) + s(A \cap B, \gamma)) + (|\beta| |\gamma| + |A \cup \varepsilon| |B \cup C|) s(A, \varepsilon)) \\&\quad - |A| |B| (|\beta| |B \cup \zeta| s(A \setminus C, C) + (|\beta| |\gamma| + |A \cup \varepsilon| |B \setminus C|) s(\alpha, C) \\&\quad + |A \cup \varepsilon| |C| (s(\alpha, \zeta \cup \gamma) + s(\alpha, B \cap C)) + (|\alpha| |\beta| + |A \cup B| |\delta \cup C|) s(\delta, C)) \\&= |B| |C| (|\delta \cup C| t(A, B \setminus A, \gamma) + |\alpha| t(A, \beta, \gamma) + |B \cup \zeta| t(A, \beta, C \setminus A)) \\&\quad + |A| |B| (|A \cup \varepsilon| t(\alpha, B \setminus C, C) + |\gamma| t(\alpha, \beta, C) + |B \cup \zeta| t(A \setminus C, \beta, C)) \\&\quad + |A| |C| ((|A \cup C| + |\zeta|) t(\alpha, B, \gamma) + |B| t(\alpha, \theta, \gamma)) \\&\quad + |A| |C| (|A \cup \varepsilon| t(\alpha, B, \zeta) + |C \cup \delta| t(\zeta, B, \gamma)) \\&\quad + 2 |A| |C| (|\delta \cup C| |A \cup \varepsilon| + |\beta| |A \cup C|) s(B, \zeta) \\&\quad + 2 |A| |B| |C| |B \cup \zeta| s(\beta, A \cap C) \geq 0.\end{aligned}$$

where the equality holds if all terms of s and t are zero.

Appendix B. Triangle Inequalities of Example 6

The triangle inequality for (4.6) can be proved as follows: Let $x = e^{p\lambda}$ and let

$$\tau(x) = \left(\frac{x|A\Delta B| + |A \cap B|}{|A \cup B|} \right) \left(\frac{x|B\Delta C| + |B \cap C|}{|B \cup C|} \right) - \left(\frac{x|A\Delta C| + |A \cap C|}{|A \cup C|} \right).$$

Then the triangle inequality for $p > 0$ is equivalent to $\tau(x) \geq 0$ for $x > 1$. The first derivative of $\tau(x)$ with respect to x is

$$\tau'(x) = 2(x-1)j(A, B)j(B, C) + j(A, B) + j(B, C) - j(A, C),$$

where j is the Jaccard distance (2.4). Since $\tau(1) = 0$ and $\tau'(x) \geq 0$ for $x \geq 1$, we have $\tau(e^{p\lambda}) \geq 0$ for $p > 0$.

The triangle inequality for (4.7) can be proved as follows: Let $y = 1 - e^{p\lambda}$ and let

$$\begin{aligned} \tau(y) &= \left(1 - \frac{|A \setminus C|}{|A \cup C|} y \right) \left(1 - \frac{|C \setminus A|}{|A \cup C|} y \right) \\ &\quad - \left(1 - \frac{|A \setminus B|}{|A \cup B|} y \right) \left(1 - \frac{|B \setminus A|}{|A \cup B|} y \right) \left(1 - \frac{|B \setminus C|}{|B \cup C|} y \right) \left(1 - \frac{|C \setminus B|}{|B \cup C|} y \right) \\ &= y\phi(y), \end{aligned}$$

where $\phi(y)$ is the cubic function of y with a negative leading coefficient. Then the triangle inequality for $p < 0$ is equivalent to $\tau(y) \geq 0$ for $y \in [0, 1]$. The function $\phi(y)$ satisfies the following inequalities: $\phi(0) \geq 0$, $\phi(1) = \tau(1) \geq 0$, $\phi'(1) \leq 0$, and $\phi''(1) \geq 0$. These inequalities can be proved by decomposition of A , B , and C into α , β , γ , δ , ε , ζ and η defined in Appendix A. Then, we have $\phi(y) \geq 0$ for $y \in [0, 1]$, therefore, $\tau(1 - e^{p\lambda}) \geq 0$ for $p < 0$.

REFERENCES

- [1] K. P. Hart, J. Nagata, and J. E. Vaughan, (eds.), *Encyclopedia of General Topology*, Elsevier, Amsterdam, 2004.
- [2] J. Nagata, *Modern General Topology*, (2nd rev. ed.), North-Holland, 1985.
- [3] W. Rucklidge, *Efficient Visual Recognition Using the Hausdorff Distance*, Lecture Notes in Computer Science, Vol. 1173, Springer, 1996.
- [4] M. M. Deza and E. Deza, *Encyclopedia of Distances*, Springer, 2009.
- [5] P. S. Bullen, *Handbook of Means and Their Inequalities*, Mathematics and Its Applications, Vol. 560, Kluwer Academic Publishers, Dordrecht, 2003.
- [6] M. O. Searcoid, *Metric Spaces*, Springer, 2007.
- [7] R. Lowen, *Approach Spaces: The Missing Link in the Topology-Uniformity-Metric Triad*, Oxford University Press, 1997.
- [8] B. S. Everitt, *Cluster Analysis*, Heinemann, London, 1980.
- [9] M. Bukatin, R. Kopperman, S. Matthews, and H. Pajoohesh, *Partial Metric Spaces*, American Mathematical Monthly, Vol. 116 (2009), No. 8, pp. 708-718.
- [10] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo Method* (2nd ed.), John Wiley & Sons, New York, 2008.
- [11] H. J. Zimmermann, *Fuzzy Set Theory and Its Applications*, (4th ed.), Kluwer Academic Publishers, 2001.

DEPARTMENT OF ARTS AND SCIENCES, OSAKA KYOIKU UNIVERSITY, 4-698-1 ASAHIGAOKA, KASHIWARA, OSAKA 582-8582, JAPAN
E-mail address: fuji@cc.osaka-kyoiku.ac.jp

METRICS BASED ON AVERAGE DISTANCE BETWEEN SETS

OSAMU FUJITA

ABSTRACT. This paper presents a distance function between sets based on an average of distances between their elements. The distance function is a metric if the sets are non-empty finite subsets of a metric space. It can be applied to produce various metric spaces on collections of sets and will be useful for analyzing complex data sets in the fields of computer science and information science. Its generalizations to include the Hausdorff metric and extensions to infinite sets for treating fuzzy sets are also discussed.

1. INTRODUCTION

A metric defined in general topology [1, 2], based on a natural notion of distance between points, is generally extensible to distance between sets or more complex elements. The Hausdorff metric is such a typical one and practically used for image data analysis [3], but it has some problems. In the Euclidean metric on \mathbb{R} , for example, the Hausdorff distance between bounded subsets of \mathbb{R} often depends only on their suprema or infima, no matter how the other elements of the sets are distributed within a certain range, which means it places importance on extremes and disregards the middle parts of the sets. This is a drawback because it is sensitive to noises, errors and outliers in analyzing real world data. There is a need to develop another metric that reflects the overall characteristics of elements of the sets.

In computer science, especially in the fields of pattern recognition, classification, information retrieval and artificial intelligence, it is important for data analysis to measure similarity or difference between data objects such as documents, images and signals. If the data objects can be represented by vectors, a conventional distance between vectors is a proper measure in their vector space. In practice, however, there are various data objects that should be dealt with in the form of collections of sets, probability distributions, graph structured data, or collections consisting of more complex data elements. To analyze these data objects, numerous distance-like functions have been developed [4], like the Mahalanobis distance and the Kullback-Leibler divergence, even though they do not necessarily satisfy symmetry and/or the triangle inequality.

As a true metric, besides the Hausdorff metric, there is another type of distance functions of sets, such as the Jaccard distance, based on the cardinality of the symmetric difference between sets or its variations. However, it measures only the size of the set difference, and takes no account of qualitative differences between individual elements. Thus, both metrics are insufficient to analyze informative data sets in which each element has its own specific meaning.

Key words and phrases. Metric, distance between sets, average distance, power mean, Hausdorff metric.

This paper presents a new distance function between sets based on an average distance. It takes all elements into account. It is a metric if the sets are non-empty finite subsets of a metric space, and includes the Jaccard distance as a special case. By using the power means [5], we obtain generalized forms that also include the Hausdorff metric. Extensions of the metric to hierarchical collections of infinite subsets will be useful for treating fuzzy sets and probability distributions.

2. PRELIMINARIES

The metric is extended to various types of generalized metrics. To avoid confusion in terminology, the following definition is used.

Definition 1. (Metric) Suppose X is a set and d is a function on $X \times X$ into \mathbb{R} . Then d is called a *metric* on X if it satisfies the following conditions, for all $a, b, c \in X$,

- M1:** $d(a, b) \geq 0$ (non-negativity),
- M2:** $d(a, a) = 0$,
- M3:** $d(a, b) = 0 \Rightarrow a = b$,
- M4:** $d(a, b) = d(b, a)$ (symmetry),
- M5:** $d(a, b) + d(b, c) \geq d(a, c)$ (triangle inequality).

The set X is called a *metric space* and denoted by (X, d) . The function d is called *distance function* or simply *distance*.

The metric is generalized by relaxing the conditions as follows:

- If d satisfies M1, M2, M4 and M5, then it is called a *pseudo-metric*.
- If d satisfies M1, M2, M3 and M5, then it is called a *quasi-metric*.
- If d satisfies M1, M2, M3 and M4, then it is called a *semi-metric*.

This terminology follows [1, 2], though the term “*semi-metric*” is sometimes referred to as a synonym of *pseudo-metric* [4].

A set-to-set distance is usually defined as follows (see, e.g., [6]): Let A and B be two non-empty subsets of X . For each $x \in X$, the distance from x to A , denoted by $\text{dist}(x, A)$, is defined by the equation

$$(2.1) \quad \text{dist}(x, A) = \inf\{d(x, a) \mid a \in A\}.$$

This is fundamental not only to the definitions of a boundary point and an open set in metric spaces but also to the generalization of a metric space to *approach space* [7]. Similarly, the distance from A to B can be straightforwardly defined by

$$(2.2) \quad \text{dist}(A, B) = \inf\{d(a, b) \mid a \in A, b \in B\}.$$

The function $\text{dist}()$ is neither a pseudo-metric nor a semi-metric. However, let $\mathcal{S}(X)$ be the collection of all non-empty closed bounded subsets of X . Then, for $A, B \in \mathcal{S}(X)$, the function $h(A, B)$ defined by

$$(2.3) \quad h(A, B) = \max\{\sup\{\text{dist}(b, A) \mid b \in B\}, \sup\{\text{dist}(a, B) \mid a \in A\}\}$$

is a metric on $\mathcal{S}(X)$, and h is called the *Hausdorff metric*. The collection $\mathcal{S}(X)$ topologized by the metric h is called a *hyperspace* in general topology.

In computer science, data sets are generally discrete and finite. A popular metric is the *Jaccard distance* (or *Tanimoto distance*, *Marczewski–Steinhaus distance* [4]) that is defined by

$$(2.4) \quad j(A, B) = \frac{|A \Delta B|}{|A \cup B|},$$

where $|A|$ is the cardinality of A , and \triangle denotes the symmetric difference: $A \triangle B = (A \setminus B) \cup (B \setminus A)$. In addition, $|A \triangle B|$ is also used as a metric.

In cluster analysis [8], the distance (2.2) is used as the *minimum distance* between data clusters for single-linkage clustering, and likewise the *maximum distance* is defined by replacing infimum with maximum for complete-linkage clustering. Moreover, the *group-average distance* (or *average distance*, *mean distance*) defined as $g(A, B)$ in the following is also typically used for hierarchical clustering. Although these three distance functions are not metrics, the group-average distance plays an important role in this paper.

Lemma 2. *Suppose (X, d) is a non-empty metric space. Let $\mathcal{S}(X)$ denote the collection of all non-empty finite subsets of X . For each A and B in $\mathcal{S}(X)$, define $g(A, B)$ on $\mathcal{S}(X) \times \mathcal{S}(X)$ to be the function*

$$(2.5) \quad g(A, B) = \frac{1}{|A||B|} \sum_{a \in A} \sum_{b \in B} d(a, b).$$

Then g satisfies the triangle inequality.

Proof. The triangle inequality for d yields $d(a, b) + d(b, c) - d(a, c) \geq 0$ for all $a, b, c \in X$. Then, for all $A, B, C \in \mathcal{S}(X)$, we have

$$(2.6) \quad \begin{aligned} & g(A, B) + g(B, C) - g(A, C) \\ &= \frac{1}{|A||B||C|} \sum_{a \in A} \sum_{b \in B} \sum_{c \in C} (d(a, b) + d(b, c) - d(a, c)) \geq 0. \end{aligned}$$

□

For ease of notation, let $s(A, B)$ be the sum of all pairwise distances between A and B such that

$$(2.7) \quad s(A, B) = \sum_{a \in A} \sum_{b \in B} d(a, b),$$

so that $g(A, B) = (|A||B|)^{-1}s(A, B)$. Since d is a metric, we have $s(A, B) \geq 0$, $s(A, B) = s(B, A)$, and $s(\{x\}, \{x\}) = 0$ for all $x \in X$. If $A = \emptyset$ or $B = \emptyset$, then $s(A, B) = 0$ due to the empty sum. If A and B are countable unions of disjoint sets, it can be decomposed as follows:

$$(2.8) \quad s\left(\bigcup_i^n A_i, \bigcup_j^m B_j\right) = \sum_i^n \sum_j^m s(A_i, B_j),$$

where $A_i \cap A_j = \emptyset = B_i \cap B_j$ for $i \neq j$. Furthermore, we define $t(A, B, C)$ by the following equation

$$(2.9) \quad t(A, B, C) = |C|s(A, B) + |A|s(B, C) - |B|s(A, C).$$

It follows from Lemma 2 that $t(A, B, C) \geq 0$ for $A, B, C \in \mathcal{S}(X)$, which is a shorthand notation of the triangle inequality (2.6).

3. METRIC BASED ON AVERAGE DISTANCE

Theorem 3. Suppose (X, d) is a non-empty metric space. Let $\mathcal{S}(X)$ denote the collection of all non-empty finite subsets of X . For each A and B in $\mathcal{S}(X)$, define $f(A, B)$ on $\mathcal{S}(X) \times \mathcal{S}(X)$ to be the function

$$(3.1) \quad f(A, B) = \frac{1}{|A \cup B| |A|} \sum_{a \in A} \sum_{b \in B \setminus A} d(a, b) + \frac{1}{|A \cup B| |B|} \sum_{a \in A \setminus B} \sum_{b \in B} d(a, b).$$

Then f is a metric on $\mathcal{S}(X)$.

Proof. The function f can be rewritten, using s in (2.7), as

$$f(A, B) = \frac{s(A, B \setminus A)}{|A \cup B| |A|} + \frac{s(A \setminus B, B)}{|A \cup B| |B|}.$$

It is non-negative and symmetric. If $A = B$, then $s(A, B \setminus A) = s(A, \emptyset) = 0$ and $s(A \setminus B, B) = s(\emptyset, B) = 0$, so that $f(A, B) = 0$. Conversely, if $f(A, B) = 0$, then $s(A, B \setminus A) = 0 = s(A \setminus B, B)$ for $A, B \in \mathcal{S}(X)$. This holds if, and only if, $B \setminus A = \emptyset = A \setminus B$, which implies $B \subseteq A$ and $A \subseteq B$. Then we have $f(A, B) = 0 \Leftrightarrow A = B$.

The triangle inequality is straightforwardly proved to be $f(A, B) + f(B, C) - f(A, C) \geq 0$ by showing that the left-hand terms are transformed into the sum of non-negative terms of s and t in (2.9). Let $A \cup B \cup C$ be decomposed into five disjoint partitions: $\alpha = A \setminus (B \cup C)$, $\beta = B \setminus (A \cup C)$, $\gamma = C \setminus (A \cup B)$, $\zeta = A \cap C \setminus B$, and $\theta = B \setminus \beta = B \cap (A \cup C)$. Then we have

$$\begin{aligned} & |A| |B| |C| |A \cup B| |B \cup C| |A \cup C| (f(A, B) + f(B, C) - f(A, C)) \\ &= |B| |C| (|\theta \cup C| t(A, B \setminus A, \gamma) + |\alpha| t(A, \beta, \gamma) + |B \cup \zeta| t(A, \beta, C \setminus A)) \\ &\quad + |A| |B| (|A \cup \theta| t(\alpha, B \setminus C, C) + |\gamma| t(\alpha, \beta, C) + |B \cup \zeta| t(A \setminus C, \beta, C)) \\ &\quad + |A| |C| (|A \setminus B| t(\alpha, B, C \setminus B) + |B| t(\alpha, \theta, \gamma) + |C \setminus B| t(A \setminus B, B, \gamma)) \\ &\quad + |A| |C| |\theta| (t(\alpha, B, \gamma) + t(\alpha, B, \zeta) + t(\zeta, B, \gamma)) \\ &\quad + 2 |A| |C| (|\theta \cup C| |A \cup \theta| + |\beta| |A \cup C|) s(B, \zeta) \\ &\quad + 2 |A| |B| |C| |B \cup \zeta| s(\beta, A \cap C) \geq 0. \end{aligned}$$

The details are given in Appendix A. \square

The function f in (3.1) can be rewritten, using g in (2.5), as

$$(3.2) \quad f(A, B) = \frac{|B \setminus A|}{|A \cup B|} g(A, B \setminus A) + \frac{|A \setminus B|}{|A \cup B|} g(A \setminus B, B).$$

In $(\mathcal{S}(X), f)$, for all $a, b \in X$, we have $f(\{a\}, \{b\}) = d(a, b)$ so that $\{\{x\} \mid x \in X\}$ is an isometric copy of X . If $A \cap B = \emptyset$, then $f(A, B) = g(A, B)$. If d is a pseudo-metric, then so is f .

Example 4. If d is the discrete metric, where $d(x, y) = 0$ if $x = y$ and $d(x, y) = 1$ otherwise, then $f(A, B)$ is equal to the Jaccard distance (2.4).

Corollary 5. Suppose (X, d) is a non-empty metric space. Let $\mathcal{S}(X)$ denote the collection of all non-empty finite subsets of X . For each A and B in $\mathcal{S}(X)$, define $e(A, B)$ on $\mathcal{S}(X) \times \mathcal{S}(X)$ to be the function

$$e(A, B) = \frac{1}{|A| |B|} \left(\sum_{a \in A} \sum_{b \in B} d(a, b) - \sum_{a \in A \cap B} \sum_{b \in A \cap B} d(a, b) \right).$$

Then e is a semi-metric on $\mathcal{S}(X)$.

Proof. Let $e(A, B)$ be rewritten as

$$e(A, B) = \frac{1}{|A||B|} (s(A \setminus B, B \setminus A) + s(A \cap B, B \setminus A) + s(A \setminus B, A \cap B)).$$

In a similar manner to the proof of Theorem 3, it can be proved that the conditions from M1 to M4, except for M5 (triangle inequality), are satisfied. \square

It is noted that the triangle inequality $e(A, B) + e(B, C) \geq e(A, C)$ holds if $|\delta||\varepsilon||\eta| = 0$, where $\delta = A \cap B \setminus C$, $\varepsilon = B \cap C \setminus A$ and $\eta = A \cap B \cap C$. Otherwise, for example, if $A = \delta \cup \eta$, $B = \delta \cup \eta \cup \varepsilon$ and $C = \eta \cup \varepsilon$ for non-empty δ , ε and η , then we have

$$\begin{aligned} & |A||B||C| (e(A, B) + e(B, C) - e(A, C)) \\ &= |\delta|s(\delta, \varepsilon) - |\varepsilon|s(\delta, \eta) - |\delta|s(\eta, \varepsilon) = -t(\delta, \eta, \varepsilon) \leq 0, \end{aligned}$$

so that the condition M5 is not generally satisfied.

4. EXTENSIONS

This section discusses future directions for generalization of the average distance based on the power mean and extensions to metrics on collections of infinite sets.

4.1. Generalization based on the power mean. The distance function (3.1) can be unified with the Hausdorff metric for finite sets by using the power mean. To simplify expressions, we use the following notation. Let $M_p^{(i)}(x \in A, \psi, w)$ be an extended weighted-power-mean of $\psi(x)$ such that

$$(4.1) \quad M_p^{(1)}(x \in A, \psi, w) = \left(\frac{1}{\sum_{x \in A} w(x)} \sum_{x \in A} w(x) (\psi(x))^p \right)^{1/p},$$

and its variation using the exponential transform of ψ ,

$$(4.2) \quad M_p^{(0)}(x \in A, \psi, w) = \frac{1}{p} \ln \left(\frac{1}{\sum_{x \in A} w(x)} \sum_{x \in A} w(x) \exp(p\psi(x)) \right),$$

where $i \in \{0, 1\}$ indicates one of the two types (4.1) and (4.2), p is an extended real number, ψ is a non-negative function of $x \in A$, and w is a weight such that $w(x) \in (0, 1]$ for each x and $\sum_{x \in A} w(x) > 0$. In addition, let $M_p^{(i)}(x \in A, \psi)$ denote the abbreviation of the equal weight case $M_p^{(i)}(x \in A, \psi, 1_A(x))$, where $1_A(x)$ is the indicator function defined by $1_A(x) = 1$ for $x \in A$ and $1_A(x) = 0$ for $x \notin A$. If there exists $x \in A$ such that $\psi(x) = 0$ for $p < 0$ in (4.1), then we define $M_p^{(1)} = 0$, which is consistent with taking the limit $\psi(x) \rightarrow 0^+$, though such a case is undefined in the conventional power mean to avoid division by zero.

The power mean includes various types of means [5], which are parameterized by p . By taking limits also for $p = 0, \pm\infty$, we have the following:

$$\begin{aligned} M_i^{(i)}(x \in A, \psi(x), 1_A(x)) &= M_i^{(i)}(x \in A, \psi(x)) = \frac{1}{|A|} \sum_{x \in A} \psi(x), \\ M_\infty^{(i)}(x \in A, \psi(x), w(x)) &= M_\infty^{(i)}(x \in A, \psi(x)) = \max\{\psi(x) \mid x \in A\}, \\ M_{-\infty}^{(i)}(x \in A, \psi(x), w(x)) &= M_{-\infty}^{(i)}(x \in A, \psi(x)) = \min\{\psi(x) \mid x \in A\}. \end{aligned}$$

Both $M_1^{(1)}$ and $M_0^{(0)}$ give the same arithmetic mean, whereas $M_0^{(1)}$ gives the geometric mean, and neither $M_\infty^{(i)}$ nor $M_{-\infty}^{(i)}$ depends on $w(x)$.

There are some forms of function composition of $M_p^{(i)}$ that include the distance function (3.1) as a special case, for example, as follows:

$$(4.3) \quad u_{p,q}^{(i,j)}(A, B) = M_p^{(i)} \left(x \in A \cup B, \left(1_{B \setminus A}(x) M_q^{(j)}(y \in A, d(x, y)) \right. \right. \\ \left. \left. + 1_{A \setminus B}(x) M_q^{(j)}(y \in B, d(x, y)) \right) \right),$$

$$(4.4) \quad v_{r,p,q}^{(k,i,j)}(A, B) \\ = M_r^{(k)} \left(S \in \{A, B\}, M_p^{(i)} \left(x \in A \cup B, 1_{S^c}(x) M_q^{(j)}(y \in S, d(x, y)) \right) \right),$$

where $i, j, k \in \{0, 1\}$ and S^c is the complement of S .

In addition, let w be extended to $w \in [0, 1]$ to include zero, though a weight for at least one summand must still be positive. Furthermore, it is assumed that $0 \cdot \infty = 0$, in order to ensure $0 \cdot 0^p = 0$ for $p < 0$ in $M_p^{(1)}$, so that the zero-weight can be used for excluding terms from averaging even if $\psi(x) = 0$ (i.e., distance zero) in the terms. Then, the function (4.3) can be simply expressed as

$$(4.5) \quad u_{p,q}^{(i,j)}(A, B) = M_p^{(i)} \left(x \in A \cup B, M_q^{(j)}(y \in A \cup B, d(x, y), w(x, y)) \right),$$

by using the weight function defined by

$$w(x, y) = [x \in A \setminus B][y \in B] + [x \in A \cap B][x = y] + [x \in B \setminus A][y \in A] \\ = 1_{A \setminus B}(x) 1_B(y) + 1_{A \cap B}(x)[x = y] + 1_{B \setminus A}(x) 1_A(y),$$

where $[\cdot]$ denotes the Iverson bracket, that is a quantity defined to be 1 whenever the statement within the brackets is true, and 0 otherwise.

The distance function f in (3.1) and the Hausdorff metric (2.3) are expressed by

$$f(A, B) = u_{i,j}^{(i,j)}(A, B) = 2 v_{k,i,j}^{(k,i,j)}(A, B), \\ h(A, B) = u_{\infty,-\infty}^{(i,j)}(A, B) = v_{\infty,\infty,-\infty}^{(k,i,j)}(A, B),$$

respectively, where $i, j, k \in \{0, 1\}$.

Although it is unclear, at present, what conditions on the parameters i, j, k, p, q , and r are necessary for (4.3) and (4.4) to be metrics, these generalized forms are capable of generating various distance functions in fact as follows:

Example 6. The exponential types $u_{p,q}^{(0,0)}(A, B)$ and $v_{r,p,q}^{(0,0,0)}(A, B)$ are written as

$$u_{p,q}^{(0,0)}(A, B) = \frac{1}{p} \ln \left(\frac{1}{|A \cup B|} \sum_{x \in A \cup B} \left(\frac{1_{B \setminus A}(x)}{|A|} \sum_{y \in A} e^{qd(x,y)} \right. \right. \\ \left. \left. + 1_{A \cap B}(x) + \frac{1_{A \setminus B}(x)}{|B|} \sum_{y \in B} e^{qd(x,y)} \right)^{p/q} \right),$$

$$v_{r,p,q}^{(0,0,0)}(A,B) = \frac{1}{r} \ln \left(\frac{1}{2} \left(\frac{1}{|A \cup B|} \sum_{x \in B \setminus A} \left(\frac{1}{|A|} \sum_{y \in A} e^{qd(x,y)} \right)^{p/q} + \frac{|A|}{|A \cup B|} \right)^{r/p} \right. \\ \left. + \frac{1}{2} \left(\frac{1}{|A \cup B|} \sum_{x \in A \setminus B} \left(\frac{1}{|B|} \sum_{y \in B} e^{qd(x,y)} \right)^{p/q} + \frac{|B|}{|A \cup B|} \right)^{r/p} \right).$$

If d is the discrete metric multiplied by a positive constant λ , then we have

$$(4.6) \quad u_{p,q}^{(0,0)}(A,B) = \frac{1}{p} \ln \left(\frac{e^{p\lambda} |A \triangle B| + |A \cap B|}{|A \cup B|} \right),$$

$$(4.7) \quad v_{0,p,q}^{(0,0,0)}(A,B) = \frac{1}{2p} \ln \left(\left(\frac{e^{p\lambda} |B \setminus A| + |A|}{|A \cup B|} \right) \left(\frac{e^{p\lambda} |A \setminus B| + |B|}{|A \cup B|} \right) \right).$$

The functions (4.6) and (4.7) are metrics for $p > 0$ and $p < 0$, respectively. The proofs of each triangle inequality are outlined in Appendix B. If $p = 0$, then it is the same situation as Example 4. By taking the limit for $p = 0$, we can see that both functions are equal to the Jaccard distance (2.4) except for the coefficient λ .

4.2. Hierarchical metric spaces. Suppose that (X, d) is a metric space and $\mathcal{S}(X)$ is the collection of all non-empty finite subsets of X . Let k be a non-negative integer, let $\mathcal{S}_{k+1}(X)$ denote the collection of all non-empty finite subsets of $\mathcal{S}_k(X)$, and let f_k be a metric on $\mathcal{S}_k(X)$, where $\mathcal{S}_1(X)$, $\mathcal{S}_0(X)$, f_1 and f_0 correspond to, respectively, $\mathcal{S}(X)$, X , f , and d in Theorem 3. For $k > 1$, in much the same way, for each \mathcal{A} and \mathcal{B} in $\mathcal{S}_k(X)$, the function $f_k(\mathcal{A}, \mathcal{B})$ can be defined by

$$(4.8) \quad f_k(\mathcal{A}, \mathcal{B}) = \frac{\sum_{A \in \mathcal{A}} \sum_{B \in \mathcal{B} \setminus \mathcal{A}} f_{k-1}(A, B)}{|\mathcal{A} \cup \mathcal{B}| |\mathcal{A}|} + \frac{\sum_{A \in \mathcal{A} \setminus \mathcal{B}} \sum_{B \in \mathcal{B}} f_{k-1}(A, B)}{|\mathcal{A} \cup \mathcal{B}| |\mathcal{B}|},$$

which generates a metric space $(\mathcal{S}_k(X), f_k)$ based on $(\mathcal{S}_{k-1}(X), f_{k-1})$. This metric will be useful for constructing hierarchical hyperspaces.

4.3. Duality. There is a kind of duality between sets and elements with respect to their distance functions. For example, we can define the functions D and d symmetrically as follows:

$$(4.9) \quad D(A, B) = |\{a \mid a \in A\} \triangle \{b \mid b \in B\}| = |A \triangle B|,$$

$$(4.10) \quad d(a, b) = |\{A \mid a \in A\} \triangle \{B \mid b \in B\}| = |\mathcal{C}(a) \triangle \mathcal{C}(b)|,$$

where $\mathcal{C}(a) = \{A \mid a \in A\}$. The set-to-set distance D in (4.9) is a metric due to the axiom of extensionality, and the element-to-element distance d in (4.10) is a pseudo-metric.

According to Theorem 3, D can be defined by f in (3.1), instead of (4.9), so that we have

$$D(A, B) = f(A, B), \\ d(a, b) = |\mathcal{C}(a) \triangle \mathcal{C}(b)|.$$

In this case, D is a pseudo-metric, depending on d , and there exist a condition that satisfy $d(a, b) = f(\cap \mathcal{C}(a), \cap \mathcal{C}(b))$. Furthermore, in the situation of Section 4.2, let (X, d) be a metric space, let $\mathcal{S}_1(X)$ be the collection of all non-empty finite subsets

of X , and let $\mathcal{C}(a) = \{A \in \mathcal{S}_1(X) \mid a \in A\}$ be an element of $\mathcal{S}_2(X)$. If d is the discrete metric, then we have

$$\begin{aligned} D(A, B) &= f_1(A, B), \\ f_2(\mathcal{C}(a), \mathcal{C}(b)) &= \kappa d(a, b), \end{aligned}$$

where κ is a certain positive real number such that $\kappa < 1$. If there exists d such that $d(a, b) = f_2(\mathcal{C}(a), \mathcal{C}(b))$, then the isometric copy of (X, d) is contained in $(\mathcal{S}_2(X), f_2)$ and d can be regarded as the function of D . In general, there possibly exist D and d that are formally expressed by

$$(4.11) \quad D(A, B) = F(\{d(a, b) \mid a \in A, b \in B\}),$$

$$(4.12) \quad d(a, b) = G(\{D(A, B) \mid a \in A, b \in B\}),$$

where F and G may be such a generalized function given in (4.5). It is interesting to consider whether F and G really exist and what features they have. In numerical analysis, D and d that are consistent with each other will be obtained by the iterative computation of (4.11) for all $A, B \in \mathcal{S}_1(X)$ and (4.12) for all $a, b \in X$, starting with an initial metric space (X, d) , if each converges to a non-trivial function.

4.4. Generalized metrics. The group average distance $g(A, B)$ in (2.5) can be regarded as a generalized metric that satisfies conditions M1 (non-negativity), M4 (symmetry) and M5 (triangle inequality) in Definition 1. In conventional topology, there has been no such generalization by dropping both conditions M2 and M3, which are usually combined together into the single axiom $d(a, b) = 0 \Leftrightarrow a = b$ (identity, reflexivity, or coincidence). Although M3 can be dropped for the pseudo-metric so as to allow $d(a, b) = 0$ for $a \neq b$, the self-distance $d(a, a) = 0$ (M2) seems to be indispensable in point-set topology where the element is a point having no size. An exception is a *partial metric* [9] that is defined to satisfy M1, M3, M4 and, instead of M5, the following partial metric triangularity,

$$(4.13) \quad d(a, b) + d(b, c) \geq d(a, c) + d(b, b).$$

In computer science or information science, the element of data sets is not merely a simple point. It may have rich contents inside. Some elements may have internal structures which cause non-zero self-distance, and some elements may have different properties each other, even though they are indiscernible from a metric point of view. The concept of distance can be used for measuring not only the difference between objects but also the cost of moving or the energy of transition between states. This is the reason why the generalization toward non-zero self-distance is worth considering. If the triangle inequality holds, it provides an upper and lower bound for them. The group average distance $g(A, B)$ can be such a typical one, and that it is simpler and more natural than the metric $f(A, B)$ in (3.1).

Incidentally, the function $g(A, B)$ is not a partial metric because it does not satisfy (4.13). On the other hand, $f(A, B)$ gives an approach to an instance of partial metrics from a special case of (4.7) in Example 6. By taking the limit as $\lambda \rightarrow \infty$ for $p < 0$ and multiplying a positive constant, for non-empty finite sets A and B , we have the following metric,

$$D_\nu(A, B) = \log |A \cup B| - \nu \log(|A| |B|),$$

where $\nu = 1/2$. Its triangle inequality is equivalent to $|A \cup B| |B \cup C| \geq |A \cup C| |B|$. This suggests, for $\nu \in [0, 1/2)$, D_ν is a partial metric on a collection of non-empty finite sets.

4.5. Extension to infinite sets. If $\mathcal{S}(X)$ is the collection of all non-null measurable subsets of (X, d) , and d is Lebesgue integrable on each element of $\mathcal{S}(X)$, then the group average distance $g(A, B)$, for $A, B \in \mathcal{S}(X)$, can be defined by

$$(4.14) \quad g(A, B) = \frac{1}{\mu(A)\mu(B)} \int_A \left(\int_B d(x, y) d\mu(y) \right) d\mu(x)$$

where $x \in A$, $y \in B$, and μ is a measure on X , and then the distance function (3.2) can be extended to

$$(4.15) \quad f(A, B) = \frac{\mu(B \setminus A)}{\mu(A \cup B)} g(A, B \setminus A) + \frac{\mu(A \setminus B)}{\mu(A \cup B)} g(A \setminus B, B).$$

If d is the discrete metric, then (4.15) is equal to the *Steinhaus distance* [4].

Example 7. Let (\mathbb{R}, d) be a metric space and let $d(x, y) = |x - y|$. For two intervals A and B , the distance function (4.15) can be expressed as

$$f(A, B) = \frac{|\sup(A) - \sup(B)| + |\inf(A) - \inf(B)|}{2} - \frac{|\sup(A) - \sup(B)| |\inf(A) - \inf(B)|}{\sup(A \cup B) - \inf(A \cup B)} [(A \subset B) \vee (A \supset B)]$$

If $A \not\subset B$ and $A \not\supset B$ (i.e., $[(A \subset B) \vee (A \supset B)] = 0$), then $f(A, B)$ is equal to the distance between the centers of A and B . This is consistent with an intuitive notion of the distance between balls in this (\mathbb{R}, d) .

If $\mathcal{S}(X)$ is the collection of all non-empty, countably infinite subsets (measure-zero sets) of X , then $g(A, B)$ and $f(A, B)$ should be defined by taking limits in (2.5) and (3.2), provided that both have definite values. In order to determine the average distance, we have to define a proper condition, which should be said to be “averageable”. The average distance will strongly depend on accumulation points in A and B , and it will require additional assumptions on the difference of the strength between the accumulation points. This requirement is closely related to a “relative measure” that is needed to obtain the ratio of the cardinality of an infinite set to the cardinality of its superset in (3.2). In conventional measure theory, however, any set of cardinality \aleph_0 is a null set having measure zero so that both counting measure and Lebesgue measure are useless for computing the ratio. It is necessary to use another measure. A feasible solution is discussed in the following section.

4.6. Estimation by sampling. In application to computational data analysis, statistical estimation by sampling is very useful for obtaining the approximate value of $f(A, B)$ when the size of the sets is very large. According to the law of large numbers, if enough sample elements are selected randomly, an average generated by those samples should approximate the average of the total population. The procedure is as follows:

- (1) Choose a superset P of $A \cup B$ as a population such that $P \supseteq A \cup B$.
- (2) Select a finite subset S of P as a sample obtained by random sampling.
- (3) Let $S_A = S \cap A$ and $S_B = S \cap B$. Then, compute $f(S_A, S_B)$ for approximation of $f(A, B)$.

The sampling process and its randomness are crucial for efficiently estimating a good approximation. Some useful hints could be found in various sampling techniques developed for Monte Carlo methods [10]. In most cases, sampling error is expected to decrease as the sample size increases, except for situations where the distribution of d has no mean (e.g., Cauchy distribution).

The notion of sampling suggests an intuitive measure to define a relative measure on a σ -algebra over a set X , which could be called “sample counting measure”. Suppose A and B are subsets of X . Let $\rho(A : B)$ be the ratio of the cardinality of A to the cardinality of B , let P be a superset of $A \cup B$, and let S_n be a non-empty finite subset of P such that $S_n = \bigcup_{i=1}^n Y_i$ where Y_i is the i -th non-empty sample randomly selected from P . Then, the ratio $\rho(A : B)$ can be determined by taking a limit of n as it approaches to ∞ as follows:

$$\rho(A : B) = \lim_{n \rightarrow \infty} \frac{|S_n \cap A|}{|S_n \cap B|},$$

if there exist such a limit and a random choice function that performs random sampling. Otherwise, instead of random sampling, systematic sampling could be available if the elements of X are supposed to be distributed with uniform density in its measurable metric space. For example, suppose there exists a finite partition of P where every part has an almost equal diameter. It seems better for S_n to have exactly one element with each of the parts.

4.7. Metrics for fuzzy sets and probability distributions. A fuzzy set can be represented by a collection of crisp sets so that the distance between fuzzy sets can be defined by the distance between the collections of such crisp sets. Let A be a fuzzy set: $A = \{(x, m_A(x)) \mid x \in X\}$, where $m_A(x)$ is a membership function, and let A_α be a crisp set called an α -level set [11] such that $A_\alpha = \{x \in X \mid m_A(x) \geq \alpha\}$. Then, A can be represented by the following set of ordered pairs: $\mathcal{C}(A) = \{(A_\alpha, \alpha) \mid \alpha \in (0, 1]\}$. The distance between two fuzzy sets A and B can be defined by $f_2(\mathcal{C}(A), \mathcal{C}(B))$, where there may be various ways to treat α . This notion is also applicable to the distance between probability distributions, where probability density functions are used instead of the membership function.

5. CONCLUDING REMARKS

We have found that, for a metric space (X, d) , there exists a distance function between non-empty finite subsets of X that is a metric based on the average distance of d . The distance function (3.1) in Theorem 3 is the most typical one, which includes the Jaccard distance as a special case where d is the discrete metric. Its extensions based on the power mean will be useful to develop generalized forms that also include the Hausdorff metric and the other various distance functions. Furthermore, the extensions to infinite subsets of X will provide metrics for measuring dissimilarity of fuzzy sets and probability distributions.

Appendix A. Triangle Inequality in Theorem 3

The triangle inequality for

$$f(A, B) = (|A \cup B| |A|)^{-1} s(A, B \setminus A) + (|A \cup B| |B|)^{-1} s(A \setminus B, B)$$

can be proved by showing the following inequality:

$$|A| |B| |C| |A \cup B| |B \cup C| |A \cup C| (f(A, B) + f(B, C) - f(A, C)) \geq 0.$$

Let $A \cup B \cup C$ be decomposed into the following seven disjoint sets:

$$\begin{aligned}\alpha &= A \setminus (B \cup C), & \beta &= B \setminus (A \cup C), & \gamma &= C \setminus (A \cup B), \\ \delta &= A \cap B \setminus C, & \varepsilon &= B \cap C \setminus A, & \zeta &= C \cap A \setminus B, & \eta &= A \cap B \cap C,\end{aligned}$$

and let $\theta = B \setminus \beta = B \cap (A \cup C)$, so that we have

$$\begin{aligned}A &= \alpha \cup \delta \cup \zeta \cup \eta, & |A| &= |\alpha| + |\delta| + |\zeta| + |\eta|, \\ B &= \beta \cup \delta \cup \varepsilon \cup \eta, & |B| &= |\beta| + |\delta| + |\varepsilon| + |\eta|, \\ C &= \gamma \cup \varepsilon \cup \zeta \cup \eta, & |C| &= |\gamma| + |\varepsilon| + |\zeta| + |\eta|, \\ \theta &= \delta \cup \varepsilon \cup \eta, & |\theta| &= |\delta| + |\varepsilon| + |\eta|.\end{aligned}$$

Taking account of (2.8) and (2.9), we have

$$\begin{aligned}& |A| |B| |C| |A \cup B| |B \cup C| |A \cup C| (f(A, B) + f(B, C) - f(A, C)) \\&= |B| |C| |B \cup C| |A \cup C| s(A, B \setminus A) + |A| |C| |B \cup C| |A \cup C| s(A \setminus B, B) \\&\quad + |A| |C| |A \cup B| |A \cup C| s(B, C \setminus B) + |A| |B| |A \cup B| |A \cup C| s(B \setminus C, C) \\&\quad - |B| |C| |A \cup B| |B \cup C| s(A, C \setminus A) - |A| |B| |A \cup B| |B \cup C| s(A \setminus C, C) \\&= |B| |C| (|\gamma| |\delta \cup C| s(A, B \setminus A) + (|\gamma| |\alpha| + |B \cup \zeta| |C \setminus A|) s(A, \beta) \\&\quad + |B \cup \zeta| |A| (s(A \setminus C, \beta) + s(A \cap C, \beta)) + (|\beta| |\gamma| + |B \cup C| |A \cup \varepsilon|) s(A, \varepsilon)) \\&\quad + |A| |C| ((|\gamma| |A \cup C| + |\zeta| |\gamma| + |\zeta| |A \cup \varepsilon|) s(\alpha, B) \\&\quad + |B| |A \cup \varepsilon| (s(\alpha, B \setminus C) + s(\alpha, B \cap C)) + |B| |\gamma| (s(\alpha, \beta) + s(\alpha, \theta)) \\&\quad + (|\delta \cup C| |\gamma| + |\delta \cup C| |A \cup \varepsilon| + |\beta| |A \cup C|) s(\zeta, B)) \\&\quad + |A| |C| ((|\alpha| |A \cup C| + |\zeta| |\alpha| + |\zeta| |\delta \cup C|) s(B, \gamma) \\&\quad + |B| |\delta \cup C| (s(B \setminus A, \gamma) + s(A \cap B, \gamma)) + |B| |\alpha| (s(\beta, \gamma) + s(\theta, \gamma)) \\&\quad + (|A \cup \varepsilon| |\alpha| + |A \cup \varepsilon| |\delta \cup C| + |\beta| |A \cup C|) s(B, \zeta)) \\&\quad + |A| |B| (|\alpha| |A \cup \varepsilon| s(B \setminus C, C) + (|\alpha| |\gamma| + |\zeta \cup B| |A \setminus C|) s(\beta, C) \\&\quad + |\zeta \cup B| |C| (s(\beta, A \cap C) + s(\beta, C \setminus A)) + (|\beta| |\alpha| + |A \cup B| |\delta \cup C|) s(\delta, C)) \\&\quad - |B| |C| (|\zeta \cup B| |\beta| s(A, C \setminus A) + (|\alpha| |\beta| + |B \setminus A| |\delta \cup C|) s(A, \gamma) \\&\quad + |A| |\delta \cup C| (s(\alpha \cup \zeta, \gamma) + s(A \cap B, \gamma)) + (|\beta| |\gamma| + |A \cup \varepsilon| |B \cup C|) s(A, \varepsilon)) \\&\quad - |A| |B| (|\beta| |B \cup \zeta| s(A \setminus C, C) + (|\beta| |\gamma| + |A \cup \varepsilon| |B \setminus C|) s(\alpha, C) \\&\quad + |A \cup \varepsilon| |C| (s(\alpha, \zeta \cup \gamma) + s(\alpha, B \cap C)) + (|\alpha| |\beta| + |A \cup B| |\delta \cup C|) s(\delta, C)) \\&= |B| |C| (|\delta \cup C| t(A, B \setminus A, \gamma) + |\alpha| t(A, \beta, \gamma) + |B \cup \zeta| t(A, \beta, C \setminus A)) \\&\quad + |A| |B| (|A \cup \varepsilon| t(\alpha, B \setminus C, C) + |\gamma| t(\alpha, \beta, C) + |B \cup \zeta| t(A \setminus C, \beta, C)) \\&\quad + |A| |C| ((|A \cup C| + |\zeta|) t(\alpha, B, \gamma) + |B| t(\alpha, \theta, \gamma)) \\&\quad + |A| |C| (|A \cup \varepsilon| t(\alpha, B, \zeta) + |C \cup \delta| t(\zeta, B, \gamma)) \\&\quad + 2 |A| |C| (|\delta \cup C| |A \cup \varepsilon| + |\beta| |A \cup C|) s(B, \zeta) \\&\quad + 2 |A| |B| |C| |B \cup \zeta| s(\beta, A \cap C) \geq 0.\end{aligned}$$

where the equality holds if all terms of s and t are zero.

Appendix B. Triangle Inequalities of Example 6

The triangle inequality for (4.6) can be proved as follows: Let $x = e^{p\lambda}$ and let

$$\tau(x) = \left(\frac{x|A\Delta B| + |A \cap B|}{|A \cup B|} \right) \left(\frac{x|B\Delta C| + |B \cap C|}{|B \cup C|} \right) - \left(\frac{x|A\Delta C| + |A \cap C|}{|A \cup C|} \right).$$

Then the triangle inequality for $p > 0$ is equivalent to $\tau(x) \geq 0$ for $x > 1$. The first derivative of $\tau(x)$ with respect to x is

$$\tau'(x) = 2(x-1)j(A, B)j(B, C) + j(A, B) + j(B, C) - j(A, C),$$

where j is the Jaccard distance (2.4). Since $\tau(1) = 0$ and $\tau'(x) \geq 0$ for $x \geq 1$, we have $\tau(e^{p\lambda}) \geq 0$ for $p > 0$.

The triangle inequality for (4.7) can be proved as follows: Let $y = 1 - e^{p\lambda}$ and let

$$\begin{aligned} \tau(y) &= \left(1 - \frac{|A \setminus C|}{|A \cup C|} y \right) \left(1 - \frac{|C \setminus A|}{|A \cup C|} y \right) \\ &\quad - \left(1 - \frac{|A \setminus B|}{|A \cup B|} y \right) \left(1 - \frac{|B \setminus A|}{|A \cup B|} y \right) \left(1 - \frac{|B \setminus C|}{|B \cup C|} y \right) \left(1 - \frac{|C \setminus B|}{|B \cup C|} y \right) \\ &= y\phi(y), \end{aligned}$$

where $\phi(y)$ is the cubic function of y with a negative leading coefficient. Then the triangle inequality for $p < 0$ is equivalent to $\tau(y) \geq 0$ for $y \in [0, 1]$. The function $\phi(y)$ satisfies the following inequalities: $\phi(0) \geq 0$, $\phi(1) = \tau(1) \geq 0$, $\phi'(1) \leq 0$, and $\phi''(1) \geq 0$. These inequalities can be proved by decomposition of A , B , and C into α , β , γ , δ , ε , ζ and η defined in Appendix A. Then, we have $\phi(y) \geq 0$ for $y \in [0, 1]$, therefore, $\tau(1 - e^{p\lambda}) \geq 0$ for $p < 0$.

REFERENCES

- [1] K. P. Hart, J. Nagata, and J. E. Vaughan, (eds.), *Encyclopedia of General Topology*, Elsevier, Amsterdam, 2004.
- [2] J. Nagata, *Modern General Topology*, (2nd rev. ed.), North-Holland, 1985.
- [3] W. Rucklidge, *Efficient Visual Recognition Using the Hausdorff Distance*, Lecture Notes in Computer Science, Vol. 1173, Springer, 1996.
- [4] M. M. Deza and E. Deza, *Encyclopedia of Distances*, Springer, 2009.
- [5] P. S. Bullen, *Handbook of Means and Their Inequalities*, Mathematics and Its Applications, Vol. 560, Kluwer Academic Publishers, Dordrecht, 2003.
- [6] M. O. Searcoid, *Metric Spaces*, Springer, 2007.
- [7] R. Lowen, *Approach Spaces: The Missing Link in the Topology-Uniformity-Metric Triad*, Oxford University Press, 1997.
- [8] B. S. Everitt, *Cluster Analysis*, Heinemann, London, 1980.
- [9] M. Bukatin, R. Kopperman, S. Matthews, and H. Pajoohesh, *Partial Metric Spaces*, American Mathematical Monthly, Vol. 116 (2009), No. 8, pp. 708-718.
- [10] R. Y. Rubinstein and D. P. Kroese, *Simulation and the Monte Carlo Method* (2nd ed.), John Wiley & Sons, New York, 2008.
- [11] H. J. Zimmermann, *Fuzzy Set Theory and Its Applications*, (4th ed.), Kluwer Academic Publishers, 2001.

DEPARTMENT OF ARTS AND SCIENCES, OSAKA KYOIKU UNIVERSITY, 4-698-1 ASAHIGAOKA, KASHIWARA, OSAKA 582-8582, JAPAN
E-mail address: fuji@cc.osaka-kyoiku.ac.jp